INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE
DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT
SDC -- A BURROUGHS COMPANY

ARPA ORDER NUMBER: 5262
PROGRAM CODE NO. NR 049-602 dated 10 August 1984 (433)
CONTRACTOR: System Development Corporation CONTRACT AMOUNT: $683,105
CONTRACT NO: N00014-85-C-0012
EFFECTIVE DATE OF CONTRACT: 4/29/85 EXPIRATION DATE OF CONTRACT: 4/29/87
PRINCIPAL INVESTIGATOR: Dr. Lynette Hirschman PHONE NO. (215) 648-7554

SHORT TITLE OF WORK: DARPA Natural Language Understanding Program

REPORTING PERIOD: 8/1/85 - 10/31/85

SDC
System Development Corporation
A Burroughs Company
P.O. Box 517, Paoli, PA 19301 Telephone (215) 648-2000

DDM-440
5 November 1985

Office of Naval Research
Department of the Navy
800 N. Quincy Street
Arlington, VA 22217-5000

Attention: Dr. Alan Meyrowitz, Code 1133

Reference: DARPA Contract No. N00014-85-C-0012

Subject: Status Report of the "DARPA Natural Language Understanding Program"
Reporting Period 8/1/85 - 10/31/85

Gentlemen:

In accordance with the referenced contract requirements System Development Corporation is pleased to submit a R & D Status Report for the DARPA Natural Language Understanding Program.

For any questions you may have please feel free to contact either Dr. Lynette Hirschman, Principal Investigator (215/648-7554) or the undersigned (215/648-2488).

Very truly yours,

SYSTEM DEVELOPMENT CORPORATION

Denise D. Mahal

Denise D. Mahal
Contract Manager

DDM/d

cc: See attached Distribution

Distribution

2 copies	Director, Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209 Attn: Program Management, Code HX1241
1 copy (cover letter only)	Defense Contract Administration Services Management Area - Philadelphia P.O. Box 7699 Phila., PA 19101-7699 Attn: Mr. A. Stein, DCASR-PHI-GAAD-E1
6 copies	Director, Naval Research Laboratory Attn: Code 2627 Washington, D.C. 20375
12 copies	Defense Technical Information Center Bldg. 5, Cameron Station Alexandria, Virginia 22314

1. Description of Progress

1.1. Grammar

The grammar has been extended to cover the first half of the CASREPS sample corpus. Both tensed and infinitival complements to verbs, adjectives and nouns are handled, as well as a wide variety of structures occurring as sentence adjuncts, such as prepositional phrases, *-ing* forms, subordinate clauses introduced by *when*, adverbs, and other structures. The restriction component of the grammar has also been augmented to enforce a number of context-sensitive constraints. For example, a restriction prevents left-recursive structures like possessives from triggering infinite structure building. As another example, since intransitive verbs in this system can take null objects, a vacuous ambiguity arises in the placement of sentence adjuncts before or after this empty object. A restriction is therefore employed to prevent assignment of a sentence adjunct to post-null object position. More generally, a class of restrictions uses look ahead to prevent the parser from attempting to apply rules that will ultimately fail. These rules have been implemented and tested, but not yet integrated into the current system.

An edit facility has been added to the grammar to dynamically edit grammar rules during parsing. This increases the efficiency of parsing by eliminating some options of grammar rules from consideration during a particular parse. Options of grammar rules are edited if a lookahead through the wordstream provides no evidence that that option is a possibility. For rules with large numbers of options, such as the object, this facility substantially increases the efficiency of parsing. The edit facility has been integrated into the current system.

An "exclusive or" operator between mutually exclusive options of grammar rules has been added. Using this operator, it is possible,for example, to prevent the system from trying to parse fragments unless there is no full sentence parse. XOR is in the current system, although the grammar does not yet make use of it.

A translator which translates bnf grammar rules to prolog is being implemented which will allow the compilation of grammar rules. This will permit translated and interpreted grammar rules to be run intermingled during development. Initial tests indicate that parsing time will be decreased approximately by a factor of four.

A Prolog version of the representation developed at NYU to mediate between the syntactic parse and semantics has been implemented and is currently being tested.

1.2. Semantics

The clause semantics interpreter has been enhanced in several ways. It is now able to check selectional restrictions on referents that have not yet been resolved by using the type information available from the sentence. That is, it can confirm that *motor* is a suitable subject for the verb *fail* in a sentence like *the motor failed* without having to know which particular motor failed. The clause semantics interpreter can also call reference resolution to resolve referents that do not occur explicitly in the sentence, such as the new replacement motor in *the motor was replaced*. These capabilities have been integrated into the system.

A treatment of nominalizations, such as *the installation of the SAC*, using Clause Semantics has been implemented and is currently being tested.

Semantic modeling of the verbs in the CASREPS and the design of the time analysis component are in progress.

1.3. Facilities

Robert Cassels, the implementer of Symbolics Prolog, came to SDC in September to discuss our problems with Prolog. As a result of this visit, SDC has become a beta test site for Symbolics

Prolog. We have installed the beta test version of Symbolics Prolog 6.1, and are porting the natural language system. A number of bugs in the previous version have been corrected in 6.1, but not all of them.

We are still awaiting receipt of Prolog on GFE. This is due in mid-November.

The version of the natural language system on the VAX has been ported from Quintus Prolog 1.0 to Quintus Prolog 1.5.

2. Change in Key Personnel

Rebecca Schiffman (Ph.D. University of Chicago, 1984) started on September 3. Rebecca Davis (B.S. Colorado State University, 1983) is working part time on Prolog for the Symbolics machines.

Karen Wieckert left SDC on August 30 to return to graduate school.

3. Summary of Substantive Information from Meetings and Conferences

3.1. SDC/NYU Meeting

SDC/NYU Meeting #5 (September 20, New York University)"

Lynette Hirschman, Martha Palmer, Deborah Dahl, John Dowding, Marcia Linebarger, and Rebecca Schiffman attended a meeting at NYU with Ralph Grishman, Ngo Thanh Nhan, and Tomasz Ksiezyk. Dahl gave a presentation on reference resolution. Nhan discussed the grammar he has implemented for the CASREPS, and Ksiezyk described his domain model of the SAC's.

3.2. Darpa Meeting

Lynette Hirschman attended the Darpa Natural Language Technology Review Panel meeting in Los Angeles at ISI (August 19-20).

3.3. Symbolics Lisp User's Group

John Dowding attended the meeting of the Mid-Atlantic Division of the Symbolics Lisp User's Group at NYU. SDC will host the next meeting in January (September 23).

4. Problems Encountered and/or Anticipated

Prolog for the Government-furnished Symbolics machine has not yet arrived. It is expected November 15.

Although a number of bugs in Symbolics Prolog have been corrected in the beta-test version of 6.1 which is currently under installation at SDC, there are various bugs remaining. In addition, problems remain with the lack of development environment and debugging facilities.

5. Action Required by the Government

Prolog for the Government-furnished Symbolics machine.

6. Fiscal Status

(1) Amount currently provided on contract:
$339,728 (funded) $683,105 (contract value)

(2) Expenditures and commitments to date:
$135,721 (through October 10, 1985)

(3) Funds required to complete work:
$204,007 (Year 1) $547,384 (Yrs. 1-2)